

Why Is Green Preservation Hard?



David S. H. Rosenthal

LOCKSS Program
Stanford University Libraries

<http://www.lockss.org/>

<http://blog.dshr.org/>

© 2013 David S. H. Rosenthal



L O T S O F C O P I E S K E E P S T U F F S A F E

Problem: Sustainable Economics



Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information

*Final Report of the Blue Ribbon Task Force on
Sustainable Digital Preservation and Access*

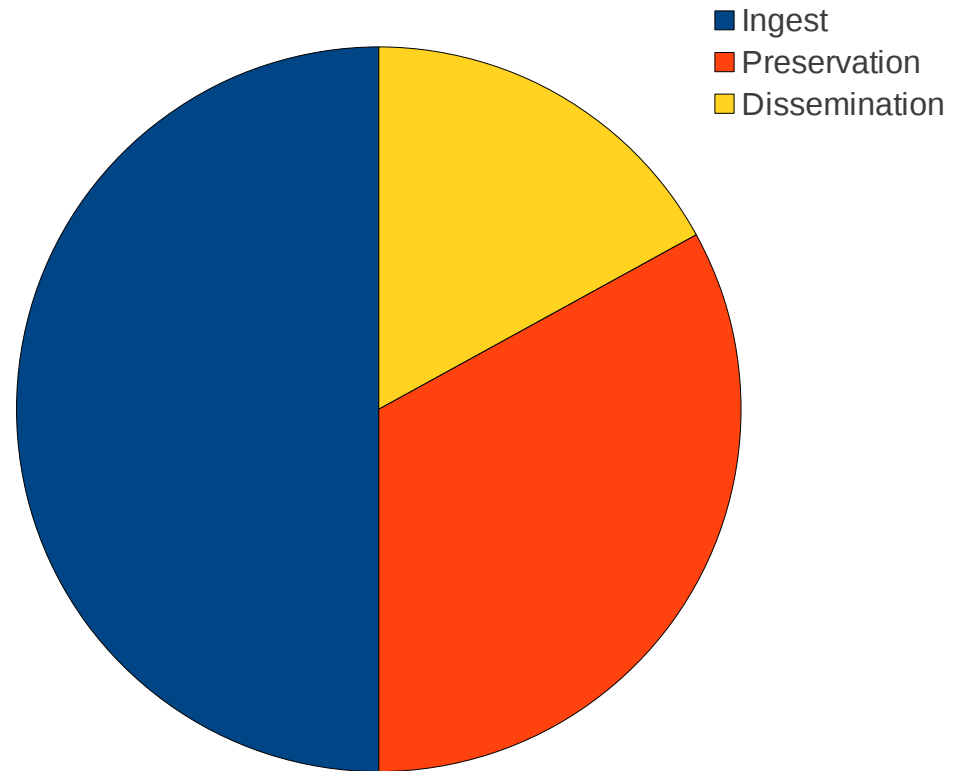
[Click to download](#)

LOTS OF COPIES KEEP STUFF SAFE

Economics of Preservation



- Blue Ribbon panel:
 - Sustainable economics
 - ... is the real problem
- Lots of research:
 - CMDP, LIFE, KRDS, PrestoPrime, ENSURE
- Rough consensus
 - 1/2 ingest
 - 1/6 dissemination
 - 1/3 preservation



LOTS OF COPIES KEEP STUFF SAFE

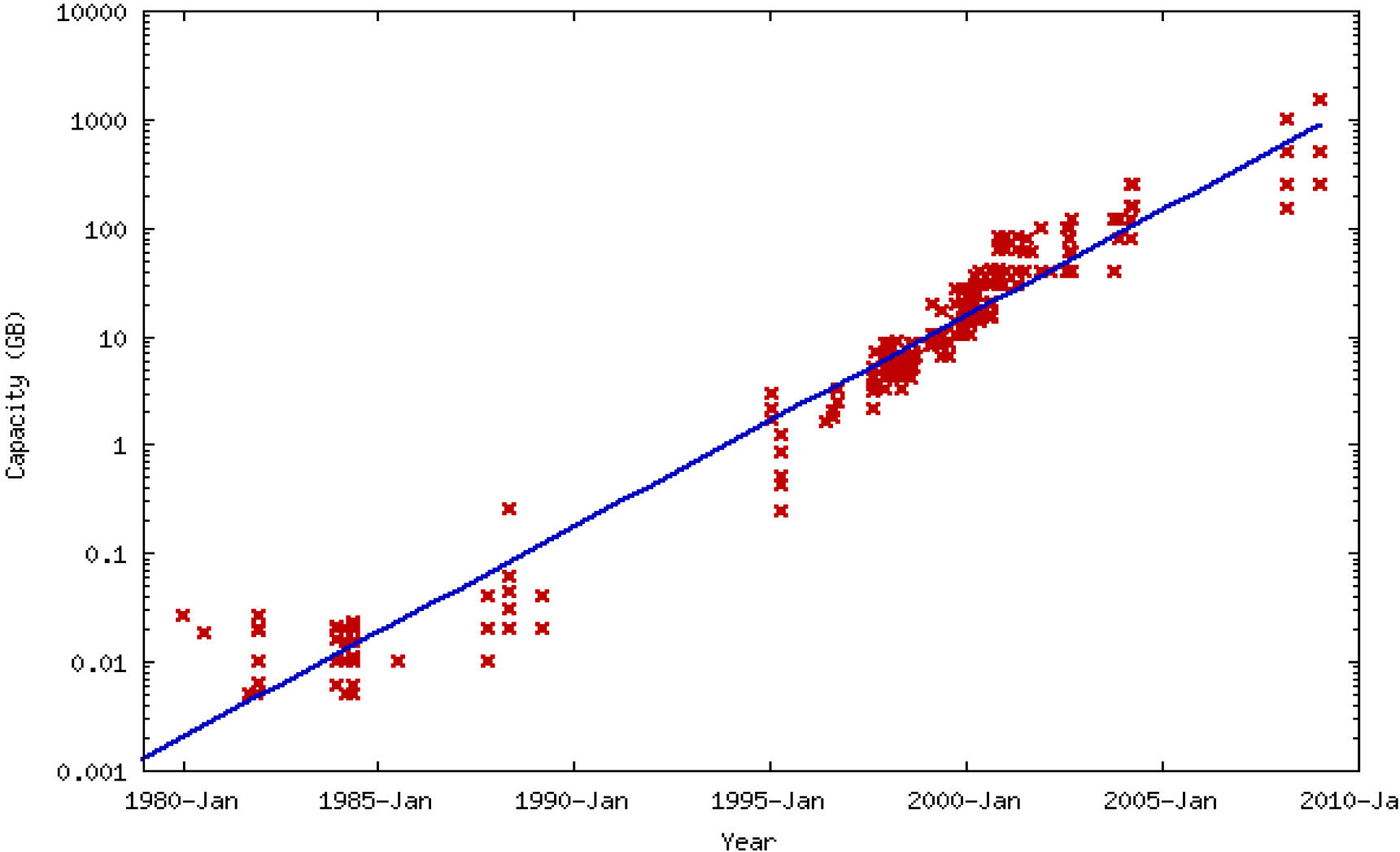
How Big A Factor Is Power?



- **Example: 4TB Seagate @ \$170 retail**
 - Operating power 7.5W, 4-yr life, 100% duty factor
 - Palo Alto Utilities: \$0.133/KWh = \$35.09 over entire life
- **Must add infrastructure cost:**
 - Backblaze 45-drive box has dual 760W supplies
 - Box uses about 1.25 times as much power as the drives.
- **Power is about 27% of total 4-yr cost:**
 - Drive: \$170 capital + \$35.09 power
 - Drive's share of box: \$43.17 capital + \$43.86 power
 - Total: \$213.17 capital + \$78.95 power



Kryder's Law



LOTS OF COPIES KEEP STUFF SAFE

Off-line Media?



- Write to long-lived off-line media, send to Iron Mtn:
 - Uses power during ingest, but not for long so not a factor
 - Risk of media and reader obsolescence
- Doesn't provide for integrity checks:
 - Integrity checks consume power
 - Handling media for checks is itself a threat
- Doesn't provide for access
 - The only way to justify preserving it in the first place

Off-line media in robots



- Robot burns power all the time:
 - Less per byte/year than disk
- Can support integrity checks, but:
 - Uses scarce bandwidth
 - Need to migrate data to newer, denser media
 - Both of which burn power
- Doesn't provide the access scholars want:
 - Large-scale data-mining from tape not feasible
 - Web access to tape not feasible

L O T S O F C O P I E S K E E P S T U F F S A F E

Key to Low-Carbon Data Centers



- Characteristics of low-carbon data centers:
 - Assign & migrate tasks to keep as few servers powered as possible
 - Key is infrastructure *agility*
- Not feasible for storage:
 - Access to preserved data has no hot-spots
 - Data-mining accesses entire collections
 - Moving a petabyte takes a long time
- Conventional data center techniques don't work
 - Preservation needs low power *and* fast access

Fast Access & Low Power



- **FAWN – Fast Array of Wimpy Nodes (C-MU):**
 - Large #s of [low-power CPU, small amount of flash]
 - Key-value queries – same speed, 2 orders of magnitude less power
- **DAWN – Durable Array of Wimpy Nodes (UCSC):**
 - Similar hardware optimized for preservation
 - Much higher capital, much lower running, much longer media life
 - With proper accounting, cheaper than hard disk
- **Caveat – proper accounting needed:**
 - Long planning horizon, proper discounting of cash flows
 - Very few organizations have such accounting

Negative Feedback



- Short-termism means you won't buy archival media:
 - So the vendors won't make it
 - They like the planned obsolescence caused by Kryder's Law
- No-one keeps disks for their working life:
 - Replaced by denser media when they no longer justify their slot
 - Even if they are still working just fine
- 2009 Seagate study of archival hard disk:
 - Not difficult technologically – bits on platter already last 2-3 decades
 - Drive mechanics need re-engineering
 - Niche product, small volumes, economics don't work

Pessimistic Conclusion



- Bulk of preserved data is going to be on hard disk:
 - At least one copy of it
 - Burning more power than it should
- Even if technology cut power by factors of 10:
 - You wouldn't buy it
 - Power isn't that big a part of the budget
 - It would be more expensive up-front; you don't care about long term
- Nevertheless, there are things you can do:
 - Won't make a big difference to cost
 - Every little helps

L O T S O F C O P I E S K E E P S T U F F S A F E